DOCUMENT RESUME

ED 269 766 CS 209 679

AUTHOR Carlman, Nancy
TITLE Variations in the Writing Performance of Grade 12 Students: Differences by Mode and Topic.
PUB DATE [85]
NOTE 21p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; Evaluation Methods; Foreign Countries; Grade 12; High Schools; *Holistic Evaluation; *Performance Factors; *Scores; Testing Problems; Test Theory; *Writing Evaluation; *Writing Research; Written Language
IDENTIFIERS Canada; *Discourse Modes; Primary Trait Scoring; Rhetorical Effectiveness; *Writing Topics

ABSTRACT

A study examined whether Canadian twelfth grade students' papers would rate differently when they were written in different modes and whether there are significant differences between global (modified holistic) scores and rhetorical effectiveness (modified primary trait) scores for the same papers. Fifty students wrote on two transactional topics and two expressive topics in counterbalanced order on four occasions in English classrooms. The papers were scored by six trained raters: three using a global scoring rubric, and three others using two rhetorical effectiveness scoring rubrics, one rubric for each mode. Analysis by "t" tests revealed significant differences between scoring methods and between modes; there were no significant differences between topics within each mode. Analysis by chi square for individuals showed no differential effects of scoring method. However, writing in different modes and on different topics within mode did produce different scores for individuals scored using rhetorical effectiveness scoring. The results suggest that although holistic-type and primary trait-type scoring methods are reliable because they produce significantly different scores on the same papers of groups of students, they should not be compared without qualification; and that writing tests for purposes such as achievement, admission, placement, or competency should require students to write on more than one topic to mitigate potential injustice to individuals. (Author/HOD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED269766

VARIATIONS IN THE WRITING PERFORMANCE OF GRADE 12 STUDENTS: DIFFERENCES BY MODE AND TOPIC

Nancy Carlman
923 E. 13th Avenue
Vancouver, BC V5T 2L7

(604) 873-0054

Sessional Lecturer
University of British Columbia

Research Assistant
Simon Fraser University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Nancy Carlman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS209679

ABSTRACT

This study investigated two questions related to the writing of grade 12 students. Are there significant differences between global (modified holistic) scores and rhetorical effectiveness (modified primary trait) scores for the same papers? Are there significant differences between scores on papers written by the same students on topics in two different modes? The analysis was done on both group scores and individuals' scores.

The study examined four 200-350 word compositions of 50 grade 12 students in Vancouver, British Columbia. They wrote on two transactional topics and two expressive topics in counterbalanced order on four occasions in English classrooms. The papers were scored by six trained raters, three rating using a global scoring rubric, three others using two rhetorical effectiveness scoring rubrics, one rubric for each mode. Interrater reliability was .73 for rhetorical effectiveness scoring and .79 for global scoring.

Analysis by *t* tests for the group produced the following results: there were significant differences between scoring methods and between modes; there were no significant differences between topics within each mode. Analysis by chi square for individuals showed no differential effects of scoring method. However, writing in different modes and on different topics within mode did produce different scores for individuals scored by rhetorical effectiveness scoring.

The following conclusions may be drawn from this study: although holistic-type and primary trait-type scoring methods are both reliable and have useful purposes for testers and teachers, because they produce significantly different scores on the same papers of groups of students, they should not be compared without qualification. If choices of topics are given on writing tests for group purposes such as large scale assessment or research, topics should all elicit the same mode and should be equated in domain of content and distance from the writer. Writing tests for purposes such as achievement, admission, placement, or competency should require students to write on more than one topic to mitigate potential injustice to individuals.

Variations in the Writing Performance of Grade 12 Students: Differences by Mode and Topic

Although most teachers and test developers are aware that individual students' writing varies in quality for many reasons including their individual responses to different writing tasks, little research has focused on these differences by themselves. There is considerable evidence that writing by the same student will vary depending on the context of the writing (Sanders & Littlefield, 1975), audience (Crowhurst & Piché, 1979; Wilkinson, Barnsley, Hanna & Swan, 1979), the student's interest and motivation (Evans, 1979; Lloyd-Jones, 1977; Newkirk, 1979; Sanders & Littlefield, 1975; Smith, 1980), familiarity with content-matter (Hilgers, 1982; Lloyd-Jones, 1977; Melton & McCready, 1982), mode (Crowhurst, 1980; Dixon & Stratta, 1981; Rosen, 1969; Spandel & Stiggins, 1980; White, 1980), and topic (Evans, 1979; Gilbert, 1980; McColly, 1970; Wiseman & Wrigley, 1958). Sometimes students just have a "bad day" (Keech, 1982).

Despite this considerable evidence and the reflections by Braddock, Lloyd-Jones, & Schoer (1963), Mellon (1975), Rosen (1969), and White (1981), many large scale tests still ask for only one sample of a student's writing (Melton & McCready, 1982). Often such one-sample tests are very important for students' futures since the tests may decide admission, scholarship, equivalency, placement, achievement, or competency. This article reports on a study of variations in individual students' writing on two similar topics in each of two modes of discourse usually considered to be widely different in the difficulty of their demands.

MODE AND TOPIC

"Mode" generally refers to descendents of Aristotle's narration, description, exposition, and argument. This system of categorizing writing is traditional in North American educational institutions, but it has several flaws as a framework for teaching and testing writing.

Mode is a description of the form of a piece of writing and may have very little to do with the aim (Kinneavy, 1971) or function (Britton, Burgess, Martin, McLeod, & Rosen, 1975) of the writing. Thus narration may be used

as an example to support an argument; or an informational piece of writing, seemingly expository, may, in fact, be attempting to persuade readers to a certain point of view.

There are also problems with these categories as far as secondary and post-secondary institutions are concerned. Narration is requested less and less frequently as students grow older. By the time they reach the later years of secondary school, they may not be asked to write anything narrative except an autobiography or in "creative writing." Description does not exist by itself except in school exercises or certain types of technical writing (Lannon, 1982; Markel, 1984).

Because there are so many problems with the conventional modes as categories, despite their persistence in composition textbooks, I chose in this study to describe the two types of writing using two of Britton et al.'s (1975) functions, expressive and transactional. Expressive writing is personal, often autobiographical, often employing narrative or descriptive forms. Transactional writing is "writing to get things done"; it conveys information and/or asserts and supports a point of view, similarly to exposition and argument.

"Topic" in this study is the writing task students are expected to respond to. Rosen (1969) defined "topic" as "the complete verbalization of the writing assignment" (p. 36). "Topic" includes both subject, for example, television or advertising or friendship, and the instructions for how the writer should handle the subject. These instructions include some or all of the following: audience, purpose, form, length, examples, and helpful suggestions, like "use specific details." Recently such writing test topics have been called "prompts" (Ruth & Murphy, 1984).

For example, the following two writing tasks ask students to write about the same subject, zoos. However, because they ask the writers to take different approaches, they are two different topics.

1) Describe a visit to a zoo.
2) Should zoos be allowed to capture and display animals from endangered species?

The purpose of this study was to discover if the same students wrote papers rated significantly differently when they were writing in different modes/functions.

METHOD

Subjects. The subjects were 50 grade 12 students randomly chosen from 6 English classes in two middle-class, urban secondary schools. Grade 12 students were chosen because their writing is likely to vary more widely than that of first year post-secondary students.

Procedure. Each student wrote four papers on four topics, one approximately every eight days, in 50 minutes English class periods. The topics were developed by the researcher and pilot tested with 70 grade 11 students to determine mode/function congruency and student interest. The order of the topics was counterbalanced, and student identities were protected through the use of blind numbers. (See Table 1)

Insert Table 1 about here.

Students' papers were typed verbatim for the purposes of the study, but the originals were graded according to teacher criteria (one teacher preferred a scale of 20, the other, percentages) and returned to the writers. Sanders and Littlefield(1975) suggest that lack of response is one problem with the rigidly controlled essay test situation.

Rating. Six graduate student raters, all former or present teachers of composition at the high school or university level, read and scored the papers using two different scoring systems; three raters used a modified holistic system, termed "global" for the purposes of the study; three others used a modified primary trait scoring system, termed "rhetorical effectiveness" for the purposes of the study. Both systems used a six point scale, and both were criterion referenced to rubrics. (See Table 2)

Insert Table 2 about here.

Raters were trained using papers written at the same time as those later randomly chosen for the analysis. Then, in two working days, they scored all 200 scripts written by the 50 students. Median interrater reliability was .79 for global scoring, .73 for rhetorical effectiveness scoring,

Table 1

The topics used in the study follow.

Transactional Topic I (T1)

> Traditional sex roles are now being questioned: divorced fathers sometimes have custody of small children; women in Edmonton do heavy, outdoor work for the Department of Public Works. Some people deplore this trend; others applaud it. State your position and support it.

Transactional Topic II (T2)

> Should a private club or other private organization have the right to limit membership on the basis of sex, ethnic group, or religion? State your position and support it.

Expressive Topic I (E1)

> All of us have to face various mental, physical, or emotional challenges as we go through life. Choose one such challenge from your life and describe it, including how you handled it and how you felt about it.

Expressive Topic II (E2)

> All of us know people who make or have made our lives more meaningful. Describe your relationship with one such person and tell why you feel that that relationship was or is so important to you.

Table 2

Scoring Rubrics for this Study

Global Scoring --for all topics.

6: This composition will be an appropriate response to the topic instructions. Although it may have occasional faults, it will be well-organized, detailed, and generally well-written.

5: This composition will be good to very good overall but may divert in some parts from the topic or not be as clearly organized as the excellent paper or have some minor mechanical errors.

4: This composition will be well-written but superficial; or it may have some errors in organization, mechanics, consistency or style which are more serious than categories 5 and 6 above.

3: This composition will misinterpret or be inconsistent in handling the topic or use overly cliched or repetitious or general language or show some errors in organization, mechanics, or style.

2: This composition will show serious weaknesses in structure, consistency, syntax, diction, and/or development.

1: This composition will exhibit very little understanding of the topic or suggest incompetence in structure, syntax, diction.

Scoring for Rhetorical Effectiveness (two rubrics)

for Transactional Topics

6: This composition takes a position, gives two or more elaborated reasons, and takes opposing positions into consideration.

5: This composition takes a position and gives two or more elaborated reasons.

4: This composition takes a position and gives one elaborated or two or more unelaborated reasons.

3: This composition takes a position and gives one unelaborated reason. Or, it

doesn't take a position but gives elaborated reasons. (Or, it discusses the examples in the topic but does not take a position--added after the raters had their training session.)

2: This composition takes a position but gives no reasons.

1: This composition does not take a clear position on the issue.

for Expressive Topics

6: This composition describes vividly and clearly and integrates this description into the personal aspects of the paper.

5: This composition describes competently and makes connections between the description and the personal aspects of the paper.

4: This composition describes competently but does not make the connection clear to the personal aspects.

3: This composition does not describe competently or does not include any personal aspects of the event.

2: This composition neither describes competently nor includes any personal aspects of the event.

1: This composition does not appear to follow the topic instructions at all.

For all three rubrics, the numbered ranks would correspond roughly to the following evaluative words: 6-excellent, 5-very good, 4-good, 3-fair, 2-poor, 1-inadequate.

calculated using Ebel's formula (1979). No adjustment was done to increase reliability because the scores were being used only for research.

Data analysis. Initially a t test was performed for the differences between means of the two scoring systems. (See Table 3)

Insert Table 3 about here.

If there had been no significant differences, further score difference tests between modes and topics could have been done using means of all six ratings. However, since there were significant differences between the ratings derived from the two scoring methods except on topic E1, analysis of group results for mode and topic was done separately for each scoring system through t tests of differences between means. Analysis of variance was not used because the numbers of scores were not equal: for each mode, the means were calculated for 100 scores; for each topic, the means were calculated for 50 scores. Individual score relationships were analyzed by chi square. Confidence levels were set at .05.

RESULTS

Differences between mode/functions. For the 50 students who wrote papers using the two different modes, there were significant differences between modes for both scoring methods. (See Table 4)

Insert Table 4 about here.

Differences between topics. To make sure that the significant differences were not associated with any particular topics, t tests were calculated for differences between topics within each mode. For neither expressive not transactional papers were there significant differences between topics for either scoring method. (See Table 5).

Insert Table 5 about here.

Table 3

Differences between Global (G) and Rhetorical Effectiveness (R) Scoring

TOPICS		Means	SD	r	t	df
T1	G	3.29	1.05	.577	-4.51*	49
	R	3.86	0.85			
T2	G	3.32	1.0	.678	-6.46*	49
	R	4.01	0.86			
E1	G	3.96	0.96	.764	-1.82	49
	R	4.12	0.73			
E2	G	3.78	1.06	.704	-3.53*	49
	R	4.17	0.94			

Note: Maximum score=6; T=transactional; E=expressive.

* $p =< .001$

Table 4

Differences between Modes for
Global and Rhetorical Effectiveness Scores

	Modes	Mean	SD	r	t	df
Global Scores	T Score	3.31	0.89	.521	-4.64**	49
	E Score	3.87	0.87			
Rhetorical Effectiveness Scores	T Score	3.94	0.64	.461	-2.16*	49
	E Score	4.12	0.66			

Note: T=transactional; E=expressive.

* p =<.05
** p =<.001

Table 5

Differences between Topics
Within Mode for Each Scoring Method

Topics	Mean	SD	r	t	df
Global Scores					
T1	3.29	1.05	.500	-0.18	49
T2	3.32	1.0			
E1	3.96	0.85	.474	1.22	49
E2	3.78	0.86			
Rhetorical Effectiveness Scores					
T1	3.86	0.96	.117	-0.95	49
T2	4.01	1.06			
E1	4.12	0.10	.256	-0.32	49
E2	4.17	0.13			

Note: T=transactional; E=expressive.

Individuals' scores. Inspection of the scores of individuals showed some students whose mean scores on two papers written on topics in the same mode differed as much as two points on a six point scale. (For an example of two such papers, see the Appendix.) Therefore, for chi square analysis of individual score relationships, a pass/fail distinction was set between scores 3 and 4; thus, scores 1-3 indicated "fail" and scores 4-6 "pass." The chi square analysis showed that the topics were significantly related within both modes for global scoring but not for rhetorical effectiveness scoring. Even for global scoring, however, a third of the students would have passed if they had written on one topic and failed if they had chosen or been assigned the other. (See Table 6)

Insert Table 6 about here.

DISCUSSION

There were significant differences between the scores on papers written in two different modes for this group of grade 12 students. Therefore, if the student chose the "wrong" mode on a writing test and if passing the test were important for some educational milestone, he or she might fail.

Even if test developers attempt to keep mode consistent and to equate topics in other ways as well, however, it is clear from inspection of individuals' scores and papers that there are quality differences for individuals that have the potential to cause them to pass or fail a writing test depending on which topic is assigned or chosen.

If it is necessary to have writing tests--and certainly it is far better to have tests which include writing samples--then writing researchers and instructors must do all they can to mitigate injustice that may be caused by the use of single topics which do not and cannot permit all students to do their best writing. Perhaps the answer will be in the direction of the California State University and Colleges Freshman English Equivalency Examination (White, 1980) which asks each student to write on two different types of topics for the same test. Perhaps the answer will be an

Table 6

Relationships between Topics within Mode and Scoring Method for Individuals

Global

Topics		T2 F	T2 P
T1	P	8	12
	F	22	8

GT χ^2=5.5**

Rhetorical Effectiveness

		T2 F	T2 P
T1	P	6	26
	F	4	14

RT χ^2=.09

Topics		E2 F	E2 P
E1	P	10	22
	F	11	7

GE χ^2=4.22*

		E2 F	E2 P
E1	P	9	31
	F	3	7

RE χ^2=.25

Note. P=pass (scores 4-6); F=fail (scores 1-3)

* p =<.05

** p =<.02

institutionalization of a folio evaluation of essays (Camp & Belanoff, 1985; Evans, 1985).

In any case, the evidence from this study indicates that instructors and test administrators cannot justify one-sample writing tests, no matter how well the topic or prompt is developed, pilot tested, or scored.

APPENDIX

Transactional Topics of Student 329

Topic T1

We are born equal. Traditional sex roles were a real fuss made up by the male chauvinists. I believe that women as as capable as men in doing their job or even better. Women are more responsible and understanding while men are aggressive and reckless in nature.

People used to say that a girl should stay at home and be pretty enough to marry a rich guy. These, in my own opinion, are all junks. We are born with the same mentality, quality and rank. For example: Queen Elizabeth I, she has all the quality of a ruler and at the sametime being a woman. The women came out to help during war world II and look at the business ladies, aren't they as capable as men? It's hard for some people to accept the fact that women are as capable as men because they are naturally brought up with the propagated idea that women are of lower rank and only best suited to be a housewife and have lot of babies.

I stated that women are more responsible and understanding because they really are. They can take care of children, work and at the same time be a housewife. As for men, they are really nothing but immature brats who only know how to fuss about when they get home and expect everything to be ready for them.

I believe that women are as good as men in that both of them have the potentiality to do what they are gifted in. Women in Edmonton or anywhere else do heavy, outdoor work for the department of public works are women who know what they are doing. As long as they are honest and sincere, and think that they have the capability, I don't see why people has to make such a fuss about it. There is rights and wrongs for everything but it is up to each and every individuals to make their own decision about it.

Sum of Global Scores of three raters: 6

Sum of Rhetorical Effectiveness Scores of three raters: 8

Topic T2

Many private clubs and organizations try to limit the membership of their organization on the basis of sex, ethnic group or religion. Exclusion of the opposite sex is common and based on clubs that provide services for only one sex. Ethnic groups may be excluded for reasons of ethnic differences between members of different ethnic groups. Religious groups may be excluded if a group preaches a religion that clashes with the religious beleifs of the club. Clubs should have the right to limit memberships on

these basis, as long as the club is of lawful intent.

Mens or womens clubs may exclude members of the opposite sex for several reasons. The club may provide activities for only one sex, and may only support facilities for only one sex. To admit members of the opposite sex would increase costs greatly. The club may also have been created as a meeting place where members of the same sex could meet and talk with others on a close social basis. If the meetings were mixed, it could cause tensions and destroy the atmosphere of the club. Membership may be limited on a sex basis due to the possible increased costs involved and also if unlimited membership would destroy the traditional atmosphere of the organization.

Ethnic groups may be excluded from clubs because their ethnic beliefs and traditions may clash with current club members. If a club is composed of a group with similar views on ethnic ideals, a club should not be forced to admit a member with ethnic ideals that clash with the rest of the club. Certain ethnic groups may also have long standing disagreements with other groups. These groups should not be allowed to mix if the organizations do not want to. Exclusion of ethnic groups should be allowed to keep the nature of the group homogenous.

Religious groups may be excluded on the basis that their religion disagrees with that of club members. A club with similar religious views should not be forced to admit members of other religions which do not agree with their religious beliefs. Many churches are exclusive for their religion and clubs should be permitted to the same.

Any club should have the right to limit their membership by sex, ethnic group, or religion as long as the group is lawful in intent. Admission of excluded members could cause increased costs, disagreements between members, and an overall loss of traditional atmosphere of the club. Private clubs should be allowed to limit membership, otherwise they are not really private clubs.

Sum of Global Scores of three raters: 16

Sum of Rhetorical Effectiveness Scores of three raters: 15

References

Braddock, R., Lloyd-Jones, R., & Schoer, L. (1963) Research in written composition. Champaign, IL: National Council of Teachers of English.

Britton, J., Burgess, T., Martin, N., McLeod, A., & Rosen, H. (1975) The development of writing abilities (11-18). London: Macmillan Educational.

Camp, R. & Belanoff, P. (1985) Models of portfolio assessment. Paper presented at the Third Annual National Testing Network in Writing Conference in Writing Assessment, San Francisco, March.

Crowhurst, M. (1980) Syntactic complexity in narration and argument at three grade levels. Canadian Journal of Education, 5(1), 6-13.

Crowhurst, M. & Piche, G. (1979) Audience and mode of discourse effects on the syntactic complexity of writing at two grade levels. Research in the Teaching of English, 13(2), 101-110.

Dixon, J., & Stratta, L. (1981) Criteria for writing in English. Southamption: Southern Regional Examinations Board.

Ebel, R. L. (1979) Essentials of educational measurement. Englewood Cliffs, NJ: Prentice-Hall.

Evans, P. J. A. (1985) Canadian activity in large scale assessment of writing. The English Quarterly, 18(4), 23-27.

Evans, P. J. A. (1979) Evaluation of writing in Ontario: Grades 8, 12, and 13. Toronto: Ontario Ministry of Education.

Gilbert, P. I. F. (1980) The effects of a primary trait scoring guide on the reliability, validity, and time used in teacher evaluation of student writing. Unpublished doctoral dissertation, Ohio State University.

Hilgers, T. L. (1982) Experimental control and the writing stimulus: The problem of unequal familiarity with content. Research in the Teaching of English, 16(3), 79-88.

Keech, C. (1982) Unexpected directions of change in students' writing performance. Paper presented at the annual conference of the American Educational Research Association, New York, March. (ERIC Document Reproduction Service No. ED 220 538)

Kinneavy, J. L. (1971) A theory of discourse. Englewood Cliffs, NJ: Prentice-Hall.

Lannon, J. M. (1982) Technical writing. (2nd edition) Boston: Little, Brown and Company.

Lloyd-Jones, R. (1977) Primary trait scoring. In C. R. Cooper & L. Odell (eds.), Evaluating writing. Urbana, IL: National Council of Teachers of English.

Markel, M. M. (1984) Technical writing. New York: St. Martin's Press.

McColly, W. (1970) What does educational research say about the judging of writing ability? Journal of Educational Research, 64(4), 148-160.

Mellon, J. (1975) National assessment and the teaching of English. Urbana, IL: National Council of Teachers of English.

Melton, V. S. & McCready, M. A. (1982) Survey of national trends in writing assessment. Ruston, LA: Louisiana Tech University.

Newkirk, T. (1979) How competent are the writing competency tests? Paper presented at the Alabama Symposium on English and American Literature, Mobile, October. (ERIC Document Reproduction Service No. ED 179 958)

Rosen, H. (1969) An investigation of the effects of differentiated writing assignments on the performance in English composition of a selected group of 15/16 year olds. Unpublished Ph. E. thesis, London University Institute of Education.

Ruth, L. & Murphy, S. (1984) Designing topics for writing assessment: Problems of meaning. College Composition and Communication, 35, 410-422.

Sanders, S. E. & Littlefield, J. H. (1975) Perhaps test essays can reflect significant improvement in freshman composition: Report on a successful attempt. Research in the Teaching of English, 9, 145-153.

Smith, M. J. (1980) Aspects of verbal behavior in written composition: Contrasts between formal and nonformal operational thinkers of college age. Unpublished coctoral dissertation, Rutgers.

Spandel, V. & Stiggins, R. (1980) Direct measures of writing skill. Portland, OR: Northwest Regional Education Laboratory.

White, E. M. (1981) Personal communication, June 25.

White, E. M. (1980) Comparison and contrast: The 1979 California State University and Colleges freshman English equivalency examination. Los Angeles: Trustees of the California State University and Colleges.

Wilkinson, A., Barnsley, G., Hanna, P., & Swan, M. (1979) Assessing language development: The Crediton Project. Paper presented at the annual conference of the Canadian Council of Teachers of English, Ottawa, May.

Wiseman, S. & Wrigley, J. (1958) Essay-reliability: The effect of choice of essay title. Educational and Psychological Measurement, 18, 129-138.